

Optimal simplex finite-element approximations of arbitrary order in curved domains circumventing the isoparametric technique

Vitoriano Ruas^{1,2*}

¹ Sorbonne Universités, UPMC Univ Paris 06, UMR 7190, Institut Jean Le Rond d'Alembert, France

² CNRS, UMR 7190, Institut Jean Le Rond d'Alembert, F-75005 Paris, France.

e-mail: vitoriano.ruas@upmc.fr

Abstract

Since the 1960's the finite element method emerged as a powerful tool for the numerical simulation of countless physical phenomena or processes in applied sciences. One of the reasons for this undeniable success is the great versatility of the finite-element approach to deal with different types of geometries. This is particularly true of problems posed in curved domains of arbitrary shape. In this case method's isoparametric version for meshes consisting of curved triangles or tetrahedra has been mostly employed to recover the optimal approximation properties known to hold for standard straight elements in the case of polygonal or polyhedral domains. However, besides obvious geometric inconveniences, the isoparametric technique helplessly requires the manipulation of rational functions and consequent compulsory use of numerical integration. The purpose of this paper is to propose, study and test a simple alternative that bypasses all the above drawbacks, without eroding qualitative approximation properties. More specifically this technique can do without curved elements and is based only on polynomial algebra.

1 Study framework

This work deals with a finite element method for solving boundary value problem posed in a two- or three-dimensional domain, with a smooth curved boundary of arbitrary shape. The principle it is based upon is close to the technique called *interpolated boundary conditions* studied in [4] for two-dimensional problems. Although the latter technique is very intuitive and has been known since the seventies (cf. [15]), it has been of limited use so far. Among the reasons for this we could quote its difficult implementation, the lack of an extension to three-dimensional problems, and most of all, restrictions on the choice of boundary nodal points to reach optimal convergence rates. In contrast our method is simple to implement in both in two- and three-dimensional geometries. Moreover optimality is attained very naturally in both cases for various choices of boundary nodal points.

In order to allow an easier description of our methodology we consider as a model the Poisson equation with Dirichlet boundary conditions, solved by different N -simplex based methods, incorporating degrees of freedom other than function values at the mesh vertices. For instance, if standard quadratic Lagrange finite elements are employed, it is well-known that approximations of an order not greater than 1.5 in the energy norm are generated (cf. [7]), in contrast to the second order ones that apply to the case of a polygonal or polyhedral domain, assuming that the solution is sufficiently smooth. If we are to recover the optimal second order approximation property something different has to be done. Since long the isoparametric version of the finite element method for meshes consisting of curved triangles or tetrahedra (cf. [17]), has been considered as the ideal way to achieve this. It turns out that, besides a more elaborated description of the mesh, the isoparametric technique inevitably leads to the integration of rational functions to compute the system matrix, which raises the delicate question on how to

*This work was partially supported by CNPq, the National Research Council of Brazil

choose the right numerical quadrature formula in the master element. In contrast, in the technique to be introduced in this paper exact numerical integration can always be used for this purpose, since we only have to deal with polynomial integrands. Moreover the element geometry remains the same as in the case of polygonal or polyhedral domains. It is noteworthy that both advantages are conjugated with the fact that no erosion of qualitative approximation properties results from the application of our technique, as compared to the equivalent isoparametric one. We should also emphasize that this approach is particularly handy, whenever the finite element method under consideration has normal components or normal derivatives as degrees of freedom. Indeed in this case the definition of isoparametric finite element analogs is not always so clear or straightforward (see. e.g. [3]).

An outline of the paper is as follows. In Section 2 we present our method to solve the model problem with Dirichlet boundary conditions in a smooth curved two-dimensional domain with conforming Lagrange finite elements based on meshes with straight triangles, in connection with the standard Galerkin formulation. Corresponding well-posedness results are demonstrated. In Section 3 we prove error estimates for the method introduced in the previous section. In Section 4 we assess the approximation properties of the method studied in the previous section by solving some two-dimensional test-problems with piecewise quadratic functions. We conclude in Section 5 with some comments on possible extensions of the methodology studied in this work. In particular we briefly show that the technique addressed in Sections 2 and 3 apply with no particular difficulty to the case of boundary value problems posed in curved three-dimensional domains.

2 Method description

Let us consider as a model the Poisson equation with Dirichlet boundary conditions in an N -dimensional smooth domain Ω with boundary Γ , for $N = 2$ or $N = 3$, namely:

$$\begin{cases} -\Delta u = f & \text{in } \Omega \\ u = d & \text{on } \Gamma, \end{cases} \quad (1)$$

where f and d are given functions defined in Ω and on Γ , having suitable regularity properties. We shall be dealing with approximation methods of order k for $k > 1$ in the standard energy norm $\|\mathbf{grad}(\cdot)\|_0$, as long as $u \in H^{k+1}(\Omega)$, where $\|\cdot\|_0$ equals $[\int_{\Omega}(\cdot)^2]^{1/2}$, i.e. it denotes the standard norm of $L^2(\Omega)$. Accordingly, we shall assume that $f \in H^{k-1}(\Omega)$ and $d \in H^{k+1/2}(\Gamma)$ (cf. [1]). Although the method to be described below applies to any d , for the sake of simplicity henceforth we shall take $d \equiv 0$.

In order to simplify the presentation here we confine the description of our method to the two-dimensional case, leaving an overview of the three-dimensional case for Section 5.

Let us be given a partition \mathcal{T}_h of Ω into straight triangles satisfying the usual compatibility conditions (see e.g. [7]). Every element of \mathcal{T}_h is considered to be a closed set and is assumed to belong to a uniformly regular family of partitions. Let $\Omega_h := \cup_{T \in \mathcal{T}_h} T$, Γ_h be the boundary of Ω_h , and h_T be the diameter of $T \in \mathcal{T}_h$. As usual we set $h := \max_{T \in \mathcal{T}_h} h_T$. Clearly enough if Ω is convex Ω_h is a proper subset of Ω . We make the more than reasonable assumptions on the mesh that no element in \mathcal{T}_h has more than one edge on Γ_h .

We also need some definitions regarding the skin $(\Omega \setminus \Omega_h) \cup (\Omega_h \setminus \Omega)$. First of all, in order to avoid non essential technicalities, we assume that the mesh is constructed in such a way that convex and concave portions of Γ correspond to convex and concave portions of Γ_h . This property is guaranteed if the points separating such portions of Γ are vertices of polygon Ω_h . In doing so, let \mathcal{S}_h be the subset of \mathcal{T}_h consisting of triangles having one edge on Γ_h . Now $\forall T \in \mathcal{S}_h$ we denote by Δ_T the set delimited by Γ and the edge e_T of T whose end-points belong to Γ and set $T' := T \cup \Delta_T$ if Δ_T is not a subset of T and $T' := \overline{T} \setminus \Delta_T$ otherwise (see Figure 1). Notice that if e_T lies on a convex portion of Γ_h , T is a proper subset of T' , while the opposite occurs if e_T lies on a concave portion of Γ_h . With such a definition we can assert that there is a partition \mathcal{T}'_h of Ω associated with \mathcal{T}_h consisting of non overlapping sets T' for $T \in \mathcal{S}_h$, besides the elements in $\mathcal{T}_h \setminus \mathcal{S}_h$.

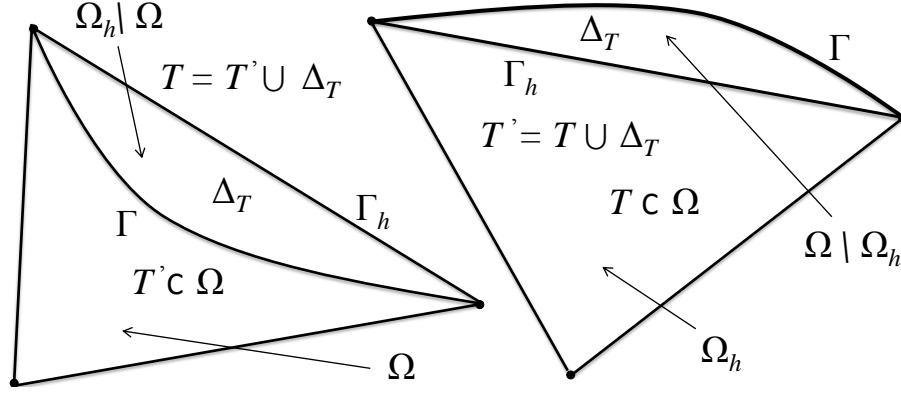


Figure 1: Skin Δ_T related to a mesh triangle T next to a convex (right) or a concave (left) portion of Γ

Next we introduce two spaces V_h and W_h associated with \mathcal{T}_h . V_h is the standard Lagrange finite element space consisting of continuous functions v defined in Ω_h that vanish on Γ_h , whose restriction to every $T \in \mathcal{T}_h$ is a polynomial of degree less than or equal to k for $k \geq 2$. For convenience we extend by zero every function $v \in V_h$ to $\Omega \setminus \Omega_h$. W_h in turn is the space of functions defined in Ω_h having the properties listed below.

1. The restriction of $w \in W_h$ to every $T \in \mathcal{T}_h$ is a polynomial of degree less than or equal to k ;
2. Every $w \in W_h$ is continuous in Ω_h and vanishes at the vertices of Γ_h ;
3. A function $w \in W_h$ is also defined in $\Omega \setminus \Omega_h$ in such a way that its polynomial expression in $T \in \mathcal{S}_h$ also applies to points in Δ_T ;
4. $\forall T \in \mathcal{S}_h, w(P) = 0$ for every P among the $k - 1$ intersections with Γ of the line passing through the vertex O_T of T not belonging to Γ and the points M different from vertices of T subdividing the edge opposite to O_T into k segments of equal length (cf. Figure 2).

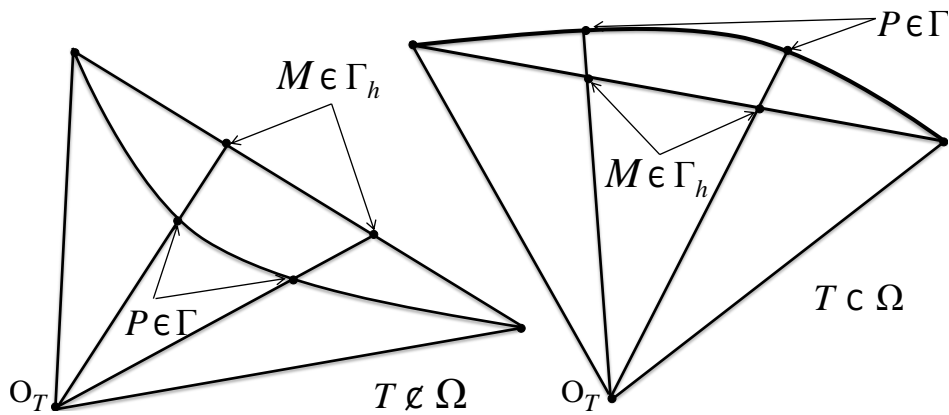


Figure 2: Construction of nodes $P \in \Gamma$ for space W_h related to lagrangian nodes $M \in \Gamma_h$ for $k = 3$

Remark 1 The construction of the nodes associated with W_h located on Γ advocated in item 4 is not mandatory. Notice that it differs from the intuitive construction of such nodes lying on normals to edges of Γ_h commonly used in the isoparametric technique. The main advantage of this proposal is an

easy determination of boundary node coordinates by linearity, using a supposedly available analytical expression of Γ . Nonetheless the choice of boundary nodes ensuring our method's optimality is really wide, in contrast to the restrictions inherent to the interpolated boundary condition method (cf. [4]). ■

The fact that W_h is a non empty finite-dimensional space is next established.

Lemma 2.1 *Let $\mathcal{P}_k(T)$ be the space of polynomials defined in $T \in \mathcal{S}_h$ of degree less than or equal to k . Provided h is small enough $\forall T \in \mathcal{S}_h$, given a set of m_k real values b_i , $i = 1, \dots, m_k$ with $m_k = (k+1)k/2$, there exists a unique function $w_T \in \mathcal{P}_k(T)$ that vanishes at both vertices of T located on Γ and at the $k-1$ points P of Γ defined in accordance with item 4. of the above definition of W_h , and takes value b_i respectively at the m_k nodes of T not located on Γ_h , corresponding to the Lagrange family of triangular finite elements (cf. [17]).*

PROOF. If the points $P \in \Gamma \cap T'$ were replaced by the corresponding $M \in \Gamma_h \cap T$, it is clear that the result would hold true, according to the well-known properties of Lagrange finite elements. The coefficients a_i for $i = 1, 2, \dots, n_k := m_k + k + 1$ of the monomials $x^m y^n$ for $m \geq 0$, $n \geq 0$ and $m + n \leq k$ are the solution of an $n_k \times n_k$ system of linear algebraic equations of the form

$$\sum_{j=1}^{n_k} a_j x_j^{\mu_i} y_j^{\nu_i} = b_i \text{ for } i = 1, \dots, n_k,$$

where μ_i and ν_i are non negative integers fulfilling $\mu_i + \nu_i \leq k$, (x_j, y_j) being the cartesian coordinates of the j -th node corresponding to Lagrange finite elements of degree k , $1 \leq j \leq n_k$. In matrix form we have $K\mathbf{a} = \mathbf{b}$ where K is a real invertible $n_k \times n_k$ matrix, $\mathbf{a} = [a_1, a_2, \dots, a_{n_k}]^T$ and $\mathbf{b} = [b_1, b_2, \dots, b_{n_k}]^T$. If each point M is replaced by the corresponding P lying on Γ with coordinates $(\tilde{x}_j, \tilde{y}_j)$ for pertaining values of j , it is clear that new coefficients \tilde{a}_j will have to be determined in order to relocate the zero values b_i to the new boundary points. In short we have to solve a new system of linear algebraic equations $\tilde{K}\tilde{\mathbf{a}} = \mathbf{b}$, where \tilde{K} is a matrix of the same form as K except for the fact that the coordinates (x_j, y_j) of points $M \in \Gamma_h$ not belonging to Γ are replaced by the coordinates $(\tilde{x}_j, \tilde{y}_j)$ of the corresponding $P \in \Gamma$. Actually the length of the segment \overline{PM} for all pairs (P, M) is bounded above by a fixed constant C times h^2 . Then it is easy to infer that the absolute values of coefficients of the increment matrix $D_K := \tilde{K} - K$ are all bounded above by $C_K h^2$ where C_K is a constant. It follows that the spectral norm $\|E_K\|$ of the matrix $E_K = K^{-1}D_K$ is bounded above by $C_D h^2$, where C_D is another constant, and therefore, if h is small enough, $\|E_K\| < 1$. Hence the matrix $I + E_K$ is invertible and so is \tilde{K} , which proves the lemma. ■

Now let us set the problem associated with spaces V_h and W_h , whose solution is an approximation of u , that is, the solution of (1). Extending f by zero in $\Omega_h \setminus \Omega$ and still denoting the resulting function by f , we wish to solve,

$$\begin{cases} \text{Find } u_h \in W_h \text{ such that} \\ a_h(u_h, v) = F_h(v) \quad \forall v \in V_h \\ \text{where } a_h(w, v) := \int_{\Omega_h} \mathbf{grad} w \cdot \mathbf{grad} v \text{ and } F_h(v) := \int_{\Omega_h} f v. \end{cases} \quad (2)$$

For convenience henceforth we refer to the nodes in a triangle belonging to the set of $(k+2)(k+1)/2$ points used to define the space of polynomials of degree less than or equal to $k > 1$ for Lagrange finite elements, as the *lagrangian nodes* (cf. [7], [17]).

Let us denote by $\|\cdot\|_{0,h}$ the standard norm of $L^2(\Omega_h)$. We next prove:

Proposition 2.2 *Provided h is sufficiently small problem (2) has a unique solution. Moreover there exists a constant $\alpha > 0$ independent of h such that,*

$$\forall w \in W_h \neq 0, \quad \sup_{v \in V_h \setminus \{0\}} \frac{a_h(w, v)}{\|\mathbf{grad} w\|_{0,h} \|\mathbf{grad} v\|_{0,h}} \geq \alpha. \quad (3)$$

We will also need

Corollary 2.3 *Provided h is sufficiently there exists a constant $\alpha' > 0$ independent of h such that,*

$$\forall w \in W_h \neq 0, \sup_{v \in V_h \setminus \{0\}} \frac{a_h(w, v)}{\|\mathbf{grad} w\|_0 \|\mathbf{grad} v\|_0} \geq \alpha'. \blacksquare \quad (4)$$

3 Error estimates

In order to derive error estimates for problem (2) we resort to the approximation theory of non coercive linear variational problems (cf. [2], [5] and [8]). At this point it is important to recall that since $d \equiv 0$, the solution u of (1) satisfies $a(u, v) = F(v) \forall v \in H_0^1(\Omega)$, where

$$a(w, v) := \int_{\Omega} \mathbf{grad} w \cdot \mathbf{grad} v \text{ and } F(v) := \int_{\Omega} f v. \quad (5)$$

Hence, owing to the construction of V_h , if Ω is convex u also fulfills $a_h(u, v) = F_h(v) \forall v \in V_h$. In case Ω is not convex, we could also extend u by zero in $\Omega_h \setminus \Omega$, to define $a_h(u, v)$. However in this case there will be a non zero residual $a_h(u, v) - F_h(v)$ for $v \in V_h$ whose order mail erode the one the approximation method (2) is supposed to attain. Nevertheless in this case such an effect can be neutralized by means of a trick to be explained later on. For the moment let us assume that Ω is convex.

Let us denote by $|\cdot|_{l,D}$ the standard semi-norm of Sobolev space $H^l(D)$ for an integer $l > 0$ (cf. [1]), D being any bounded domain of \mathbb{R}^2 with non zero measure. We have,

Theorem 3.1 *As long as h is sufficiently small, if Ω is convex and the solution u of (1) for $d \equiv 0$ belongs to $H^{k+1}(\Omega)$, the solution u_h of (2) satisfies for $k > 1$ and a suitable constant C independent of h and u :*

$$\|\mathbf{grad}(u - u_h)\|_{0,h} \leq Ch^k |u|_{k+1,\Omega}. \blacksquare \quad (6)$$

Corollary 3.2 *As long as h is sufficiently small, if Ω is convex and the solution u of (1) for $d \equiv 0$ belongs to $H^{k+1}(\Omega)$, the solution u_h of (2) satisfies for $k > 1$ and a suitable constant C' independent of h and u :*

$$\|\mathbf{grad}(u - u_h)\|_0 \leq C' h^k |u|_{k+1,\Omega}. \blacksquare \quad (7)$$

Let us now consider the non convex case. Since both Γ and u are sufficiently smooth by assumption, we can define an equally smooth extension of u , say \bar{u} , to an open narrow strip Δ_{Ω} lying in $\mathbb{R}^2 \setminus \bar{\Omega}$, in such a way that the boundary of Δ_{Ω} is the union of Γ and the outer boundary Γ' of Δ_{Ω} with $\Gamma \cap \Gamma' = \emptyset$. Typically the distance of any point $Q \in \Gamma$ to Γ' measured along the outer normal to Γ at Q equals $C_0 h_0^2 \ll 1$, where C_0 is a constant and h_0 is an upper bound for all mesh steps h being considered in our analysis. In doing so we assert that Ω' contains all possible polygons Ω_h associated with (2). Let $\bar{f} := -\Delta \bar{u}$. This means that similarly f is extended by \bar{f} to Δ_{Ω} . In short, \bar{u} is the solution of a modification of (1) in which Ω is replaced with $\Omega' := \Omega \cup \Gamma \cup \Delta_{\Omega}$ f is replaced with \bar{f} , Γ is replaced with Γ' and d represents the trace of \bar{u} on Γ' . Notice that $\bar{u}|_{\Omega} = u$, $\bar{f}|_{\Omega} = f$ and $u = 0$ on Γ , according to our assumptions.

Now instead of solving (2) we consider that we are solving $a_h(u_h, v) = \int_{\Omega_h} \bar{f} v \forall v \in V_h$, with $u_h \in W_h$. The point here is that now the residual $a_h(\bar{u}, v) - \bar{F}_h(v)$ vanishes $\forall v \in V_h$. Furthermore we can apply the very same steps in Theorem 3.1 leading to (6) to conclude:

Theorem 3.3 *Assume that the solution u of (1) for $d \equiv 0$ belongs to $H^{k+1}(\Omega)$. Let $u_h \in W_h$ be the unique solution of the variational equation*

$$a_h(u_h, v) = \bar{F}_h(v) \forall v \in V_h \text{ where } \bar{F}_h(v) := \int_{\Omega_h} \bar{f} v,$$

with $\bar{f} = -\Delta \bar{u}$, \bar{u} being an extension of u to the above defined domain Ω' strictly containing Ω . Then for $k > 1$ and a suitable constant C independent of h and u it holds.

$$\| \mathbf{grad}(\bar{u} - u_h) \|_{0,h} \leq Ch^k |\bar{u}|_{k+1,\Omega}. \quad \blacksquare \quad (8)$$

It is noteworthy that $\| \mathbf{grad}(\bar{u} - u_h) \|_{0,h}$ tends to zero as h goes to zero. It follows that $\mathbf{grad} u_h$ tends to $\mathbf{grad} u$ in the sense of $L^2(\Omega)$ as an $O(h^k)$.

Now we observe that \bar{u} restricted to Ω also fulfills the relation $a_h(\bar{u}, v) = F_h(v) \equiv \bar{F}(v) \forall v \in V_h$ in case Ω is convex. Therefore we may apply the classical Aubin-Nitsche duality argument to derive error estimates for (2) in the L^2 -norm. We state below a would-be theorem that leads to a sub-optimal convergence result in the L^2 -norm. The complete proof of this result is very technical and yet requires some tools which to the best of author's knowledge are still unavailable.

Conjecture:

Under the same assumptions as in Theorem 3.3 for a constant C_0 independent of h and u it holds:

$$\| \bar{u} - u_h \|_{0,h} \leq C_0 h^{k+1} |\bar{u}|_{k+1,\Omega}. \quad (9)$$

The proof of (9) is not simple. Even the one of a related suboptimal estimate is not so obvious, namely,

$$\| \bar{u} - u_h \|_{0,h} \leq C_0 h^{k+1/2} |\bar{u}|_{k+1,\Omega}. \quad (10)$$

Just to illustrate this assertion we list below the main steps that could be followed to prove (10). Some of them marked with a *, though more than plausible, require either a more careful scrutiny or a detailed proof. For the sake of brevity we just sketch this would-be proof restricted to the case of a convex Ω . In the inequalities that follow C represents different constants.

- I) Write $\| \bar{u} - u_h \|_0 = \sup_{g \in L^2(\Omega) \setminus \{0\}} \frac{\int_{\Omega} (\bar{u} - u_h) g}{\| g \|_0}$;
- II) Take $v \in H^2(\Omega) \cap H_0^1(\Omega)$ as the solution of $-\Delta v = g \in L^2(\Omega)$ which satisfies $\| v \|_{2,\Omega} \leq C \| g \|_0$;
- III) Rewrite $\| \bar{u} - u_h \|_0 \leq C \sup_{v \in H^2(\Omega) \setminus \{0\}} \frac{\int_{\Omega} (u_h - \bar{u}) \Delta v}{\| v \|_{2,\Omega}}$;
- IV) Apply First Green's Identity to derive $\int_{\Omega} (u_h - \bar{u}) \Delta v = a(\bar{u} - u_h, v) + \oint_{\Gamma} u_h \partial v / \partial n$ where $\partial(\cdot) / \partial n$ denotes the outer normal derivative on Γ ;
- V*) Use the fact that $u_h(P) = 0 \forall P \in \mathcal{L}_S$ and $\forall S \in \mathcal{S}_h$ to estimate $\| u_h \|_{-1/2,\Gamma} \leq \| u_h \|_{0,\Gamma} \leq C [\sum_{T \in \mathcal{S}_h} h_T^{2(k+1)} \| u_h \|_{k+1,T' \cap \Gamma}^2]^{1/2}$, $\| \cdot \|_{s,D}$ being the standard norm of Sobolev space $H^D(\Gamma)$ for $s \in \mathbb{R}$ and $D \subseteq \Gamma$;
- VI*) From a variant of inverse inequalities in [9] infer that $\| u_h \|_{k+1,T' \cap \Gamma} \leq Ch_T^{-k-1/2} \| u_h \|_{1/2,T' \cap \Gamma}$, and consequently that $\| u_h \|_{-1/2,\Gamma} \leq Ch^{1/2} \| u_h \|_{1/2,\Gamma}$;
- VII) Apply the Trace theorem and Friedrichs-Poincaré inequality for $H^1(\Omega)/\mathbb{R}$ to bound $\| u_h \|_{-1/2,\Gamma} \leq Ch^{1/2} \| \mathbf{grad}(\bar{u} - u_h) \|_0$;
- VIII) Combine the Trace Theorem with (8) to derive $\oint_{\Gamma} u_h \partial v / \partial n \leq Ch^{k+1/2} \| \bar{u} \|_{k+1,\Omega} \| v \|_{2,\Omega}$;
- IX) Using the property $a(\bar{u} - u_h, v_h) = 0 \forall v_h \in V_h$ establish that $a(\bar{u} - u_h, v) \leq \| \mathbf{grad}(\bar{u} - u_h) \|_0 (Ch \| v \|_{2,\Omega} + \| \mathbf{grad} v \|_{0,\Omega \setminus \Omega_h})$;
- X) Apply First Green's Identity to derive $\| \mathbf{grad} v \|_{0,\Omega \setminus \Omega_h}^2 = \oint_{\Gamma_h} v \partial v / \partial n_h - \int_{\Omega \setminus \Omega_h} v \Delta v$, where $\partial(\cdot) / \partial n_h$ is the normal derivative along Γ_h directed inwards Ω_h ;
- XI*) Extend the Trace Theorem in Ω to prove that $[\oint_{\Gamma_h} (\partial v / \partial n_h)^2]^{1/2} \leq C \| v \|_{2,\Omega} \forall h$, and observe that $\| \Delta v \|_{0,\Omega \setminus \Omega_h} \leq C \| v \|_{2,\Omega}$;
- XII) $\forall Q \in \Omega \setminus \Omega_h$ write $v^2(Q) = [\int_R^Q \partial v / \partial y dy]^2$, where $R \in \Gamma$ and Q lie on the same perpendicular r to a generic edge of Γ_h , y being the abscissa along r ; conclude that $[\oint_{\Gamma_h} v^2]^{1/2}$ and $[\int_{\Omega \setminus \Omega_h} v^2]^{1/2}$ are bounded above by $Ch^\beta \| v \|_{2,\Omega}$, with $\beta = 1$ and $\beta = 2$, respectively.

Infer from the twelve items above the validity of (10). \blacksquare

J	\longrightarrow	4	8	16	32	64
$\ \mathbf{grad}(u - u_h) \ _{0,h}$	\longrightarrow	0.539250 E-2	0.143615 E-2	0.367543 E-3	0.927840 E-4	0.232998 E-4
$\ u - u_h \ _{0,h}$	\longrightarrow	0.149655 E-3	0.183918 E-4	0.230310 E-5	0.289312 E-6	0.363247 E-7
$\ u - u_h \ _{0,\infty,h}$	\longrightarrow	0.604305 E-2	0.172473 E-2	0.446493 E-3	0.112615 E-3	0.282163 E-4

Table 1: Absolute errors in different senses for Test-problem 1.

Remark 2 The estimate (10) should also hold if Ω is not convex. However in this case even more complicated technicalities come into play. That is why we refrained from developing them here. ■

Remark 3 As long as the boundary nodes are placed at strategic locations, optimal $O(h^{k+1})$ error estimates in the L^2 -norm can be proven to hold. This is a consequence of results known since long, derived by Nitsche [10] and Scott [15] among other authors. However, in contrast to our method, extensions to the three-dimensional case of these results are lacking. Moreover our estimates in Theorems 3.1 and 3.2 hold for sets of boundary nodes chosen in a rather unconstrained manner.

Remark 4 (10) is certainly not the sharpest estimate one can hope for, neither in the convex nor in the non convex case. This is illustrated by numerical examples given in the next section in which L^2 -errors in terms of h^{k+1} were observed. The non optimality of (10) seems to result from step V). As a matter of fact, if an estimate of the form $\| u_h \|_{-1/2,\Gamma} \leq C[\sum_{T \in \mathcal{S}_h} h_T^{2k+3} \| u_h \|_{k+1,T' \cap \Gamma}^2]^{1/2}$ could be used, together with the inverse inequalities $\| u_h \|_{k+(2-\eta)/2,T' \cap \Gamma} \leq Ch_T^{-1/2} \| u_h \|_{k+(1-\eta)/2,T' \cap \Gamma}$ for $\eta = 0, \dots, 2k$, one would be led to a sharper bound in terms of an $O(h^{k+1})$. However, as pointed out in [15], only an estimate of the form $\| u_h \|_{-1/2,\Gamma} \leq C[\sum_{T \in \mathcal{S}_h} h_T^{2k+3} \| u_h \|_{k+3/2,T' \cap \Gamma}^2]^{1/2}$ is guaranteed and yet for a particular choice of nodal points on Γ . This is clearly not sufficient to improve the upper bound given in V) and hence to prove (9). ■

4 Numerical experiments

In order to illustrate the error estimates derived in the previous section we solved equation (1) with our method in two test-cases, taking $k = 2$.

4.1 Test-problem 1

Here Ω is the ellipse delimited by the curve $(x/e)^2 + y^2 = 1$ with $e > 0$ for an exact solution u given by $u = (e^2 - e^2x^2 - y^2)(e^2 - x^2 - e^2y^2)$. Thus we take $f := -\Delta u$ and $d \equiv 0$ and owing to symmetry we consider only the quarter domain given by $x > 0$ and $y > 0$ by prescribing Neumann boundary conditions on $x = 0$ and $y = 0$. We take $e = 0.5$ and compute with quasi-uniform meshes defined by a single integer parameter J , constructed by the procedure proposed in [12]. Roughly speaking the mesh of the quarter domain is the polar coordinate counterpart of the standard uniform mesh of the unit square $(0, 1) \times (0, 1)$ whose edges are parallel to the coordinate axes and to the line $x = y$.

In Table 1 we display the absolute errors in the norm $\| \mathbf{grad}(\cdot) \|_{0,h}$ and in the norm of $L^2(\Omega_h)$ for increasing values of J , more precisely $J = 2^m$ for $m = 2, 3, 4, 5, 6$. We also show the evolution of the maximum absolute errors at the mesh nodes denoted by $\| u - u_h \|_{0,\infty,h}$.

As one infers from Table 1, the approximations obtained with our method perfectly conform to the theoretical estimate (6). Indeed as J increases the errors in the gradient L^2 -norm decrease roughly as $(1/J)^2$, as predicted. The error in the L^2 -norm in turn tends to decrease as $(1/J)^3$, while the maximum absolute error seem to behave like an $O(h^2)$.

I	\longrightarrow	4	8	16	32	64
$\ \mathbf{grad}(\bar{u} - u_h) \ _{0,h}$	\longrightarrow	0.132906 E-1	0.334304 E-2	0.838061 E-3	0.209734 E-3	0.524545 E-4
$\ \bar{u} - u_h \ _{0,h}$	\longrightarrow	0.400090 E-3	0.491773 E-4	0.610753 E-5	0.761759 E-6	0.951819 E-7
$\ \bar{u} - u_h \ _{0,\infty,h}$	\longrightarrow	0.923815 E-2	0.238444 E-2	0.600821 E-3	0.150500 E-3	0.376434 E-4

Table 2: Absolute errors in different senses for Test-problem 2.

4.2 Test-problem 2

The aim of this Test-problem is to assess the behavior of our method in the case where Ω is non convex. Here we solve (1) for the following data: Ω is the annulus delimited by the circles given by $r = e < 1$ and $r = 1$ with $r^2 = x^2 + y^2$, for an exact solution u given by $\bar{u} = (r - e)(1 - r)$ with $\bar{f} := -\Delta \bar{u}$ and $d \equiv 0$. Again we apply symmetry conditions on $x = 0$ and $y = 0$. We take $e = 0.5$ and compute with quasi-uniform meshes defined by two integer parameters I and J , constructed by subdividing the radial range $(0.5, 1)$ into J equal parts and the angular range $(0, \pi/4)$ into I equal parts. In this way the mesh of the quarter domain is the polar coordinate counterpart of the $I \times J$ mesh of the rectangle $(0, \pi/4) \times (0.5, 1)$ whose edges are parallel to the coordinate axes and to the line $x = \pi(y - 0.5)/2$. In Table 1 we display the absolute errors in the norm $\| \mathbf{grad}(\cdot) \|_{0,h}$ and in the norm of $L^2(\Omega_h)$ for $I = 2J$, for increasing values of I , namely $I = 2^m$ for $m = 2, 3, 4, 5, 6$. We also show the evolution of the maximum absolute errors at the mesh nodes denoted by $\| u - u_h \|_{0,\infty,h}$.

As one can observe, here again the quality of the approximations obtained with our method are in very good agreement with the theoretical result (8), for as J increases the errors in the gradient L^2 -norm decrease roughly as h^2 , as predicted. On the other hand here again the errors in the L^2 -norm tend to decrease as h^3 and the maximum absolute errors behave like an $O(h^2)$.

5 Possible extensions and conclusions

To conclude we make some comments on the methodology introduced in this work, thereby showing how universal it can be.

5.1 General considerations

The method advocated in this work to solve the Poisson equation in curved domains with classical Lagrange finite elements provides a simple and reliable manner to overcome technical difficulties brought about by more complicated problems and interpolations. This issue was illustrated in a presentation at ICNAAM 2016 (cf. [13]), where we applied our technique to a Hermite analog of the Raviart-Thomas mixed finite element method of the lowest order to solve Maxwell's equations with Neumann boundary conditions. In a forthcoming paper we intend to complete this study by extending the technique to the whole Raviart-Thomas family [11], and to present the corresponding numerical analysis.

Hermite finite element methods to solve fourth order problems in curved domains with normal derivative degrees of freedom can also be dealt with very easily by means of our method. This is also shown in a paper to appear in due course [14].

The solution of (1) with a non zero d using our method is straightforward. Indeed, obviously enough, it suffices to assign the value of d at each node belonging to the true boundary Γ for any boundary element, that is, any element having an edge contained in Γ_h . The error estimates derived in this paper trivially extends to this case as the reader can certainly figure out. On the other hand in the case of Neumann boundary conditions $\partial u / \partial n = 0$ on Γ (provided f satisfies the underlying scalar condition) our method coincides with the standard Lagrange finite element method. Incidentally we recall that in case inhomogeneous Neumann boundary conditions are prescribed optimality can only be recovered if

the linear form F_h is modified in such a way that boundary integrals for boundary elements T are shifted to the curved boundary portion of an element \tilde{T} sufficiently close to the one of the corresponding curved element T' . But this is an issue that has nothing to do with our method, which is basically aimed at resolving those related to the prescription of degrees of freedom in the case of Dirichlet boundary conditions.

As the reader has certainly noticed our method leads to well-posed problems, though with a non symmetric matrix. Moreover in order to compute the element matrix and right side vector for a boundary element (in \mathcal{S}_h), we have to determine the inverse of an $n_k \times n_k$ matrix. However this extra effort should by no means be a problem at the current state-of-the art of Scientific Computing, as compared to the situation by the time isoparametric finite elements were introduced.

Saying a few words about the extremely important three-dimensional case is mandatory, and that is how we close this work.

5.2 A short account of the three-dimensional case

For the sake of brevity we confine ourselves to the model problem (1).

First of all for $N = 3$ we make the very realistic assumption that an element $T \in \mathcal{T}_h$ has at most one face on Γ_h , and if no such a face exists T has at most one edge on Γ_h . Actually we have to consider two subsets of \mathcal{T}_h , namely the subset \mathcal{S}_h consisting of tetrahedra having one face on Γ_h and the subset \mathcal{R}_h consisting of tetrahedra having exactly one edge on Γ_h . In contrast to the two-dimensional case, for $T \in \mathcal{S}_h$ it is not possible to define the set Δ_T delimited by Γ and the face F_T of T contained in Γ_h , or equivalently the three skins associated with the three edges of F_T , in such a way that an underlying space W_h of continuous functions is generated. Otherwise stated, in the three-dimensional case we have to deal with a non conforming space W_h . However this is not really a problem since the test-function space V_h remains conforming. Nevertheless, at least from the formal point of view one had better employ a systematic way to extend or restrict the elements in \mathcal{S}_h in order to construct a companion mesh of the whole Ω consisting of non overlapping straight elements $T \in \mathcal{T}_h \setminus \mathcal{S}_h$ and curved elements T' associated with $T \in \mathcal{S}_h$. Among other possibilities we can proceed as follows. For the latter elements, T' is delimited by Γ , the boundary portions of T lying inside Ω , and three skins δ_e corresponding to the three edges of the face $F_T \subset \Gamma_h$ generically denoted by e . δ_e lies on the plane containing e that bisects the dihedral formed by two mesh faces whose intersection is e . Typically the pair of faces under consideration would correspond to the largest angle formed by two such faces. The interpolation points on Γ pertaining to $T \in \mathcal{S}_h$ which are nodal points of W_h , are simply the intersections with Γ of the perpendicular to e in δ_e passing through the lagrangian nodes of e . It is noteworthy that such nodes are interpolation nodes replacing lagrangian nodes of e for an element $T \in \mathcal{R}_h$ having e as an edge, although it is not necessary to consider any extension T' of such a T . For every boundary mesh edge e we denote by \mathcal{L}_e the set of $k + 1$ nodes belonging to $\bar{\delta}_e$ defined in the above manner.

This apparently complicated definition is aimed at ensuring that there is an extension \mathcal{T}'_h of the partition \mathcal{T}_h consisting of non overlapping sets T' extending or restricting T , or doing both things at a time (typically $T' := T \cup \Delta_T$ or $T' := \overline{T \setminus \Delta_T}$ according to the local configuration of Γ), besides the elements in $\mathcal{T}_h \setminus \mathcal{S}_h$.

Now for $w \in W_h$, $\forall T \in \mathcal{S}_h \cup \mathcal{R}_h$ and for every edge $e \subset T \cap \Gamma_h$, $w(P) = d(P)$ for all $P \in \mathcal{L}_e$. If $T \in \mathcal{R}_h$ all the remaining $(k+5)(k+1)k/6$ nodes used to define $w|_T$ for $w \in W_h$ are lagrangian nodes of T . As for $T \in \mathcal{S}_h$, besides the $3k$ nodes in the three pertaining δ_e s and its $(k+2)(k+1)k/6$ lagrangian nodes not lying on Γ_h , for $k > 2$ only, the remaining $(k-1)(k-2)/2$ nodes of $T \in \mathcal{S}_h$ associated with W_h are the intersections with Γ of the line passing through the vertex O_T of T not belonging to Γ and the points subdividing the face opposite to O_T into k^2 equal triangles, except those lying on the edges of F_T . Notice that, provided h is small enough, there no chance for two out of thus constructed $(k+3)(k+2)(k+1)/6$ nodes of $T \in \mathcal{S}_h \cup \mathcal{R}_h$ to be too close to each other, let alone to coincide.

Once the space W_h is defined in accordance with the above constructions, the approximate problem (2) can be posed in the same way as in the two-dimensional case. Corresponding existence, uniqueness

and uniform stability results can be demonstrated in basically the same manner as in Section 2. As for error estimates, qualitative results equivalent to those proved in Section 3 can be expected to hold. Nonetheless their proof is at the price of several additional technicalities, especially in the non convex case. We intend to address all those issues more thoroughly in a forthcoming paper.

References

- [1] R.A. Adams. *Sobolev Spaces*. Academic Press, N.Y., 1975.
- [2] I. Babuška. The finite element method with Lagrange multipliers. *Numerische Mathematik*, 20 (1973), 170–192.
- [3] F. Bertrand, S. Münzenmaier and G. Starke. First-order system least-squares on curved boundaries: higher-order Raviart–Thomas elements. *SIAM J. Numerical Analysis* 52-6 (2014), 3165-3180.
- [4] S.C. Brenner and L.R.Scott. *The Mathematical Theory of Finite Element Methods*. Texts in Applied Mathematics 15, Springer, 2008.
- [5] F. Brezzi. On the existence, uniqueness and approximation of saddle-point problems arising from Lagrange multipliers. *RAIRO Analyse Numérique*. 8-2 (1974), 129-151.
- [6] F. Brezzi and M. Fortin (eds.). *Mixed and Hybrid Finite Element Methods*. Springer Series in Computational Mathematics, Vol. 15, 1991.
- [7] P.G. Ciarlet. *The Finite Element Method for Elliptic Problems*. North Holland, Amsterdam, 1978.
- [8] J.A. Cuminato and V. Ruas. *Unification of distance inequalities for linear variational problems. Computational and Applied Mathematics*, 34 (2015), 1009-1033.
- [9] E.H. Georgoulis. Inverse type estimates on hp-finite element spaces and applications. *Mathematics of Computation*, 77-261 (2008), 201–219.
- [10] J. Nitsche. On Dirichlet problems using subspaces with nearly zero boundary conditions. *The Mathematical Foundations of the Finite Element Method with Applications to Partial Differential Equations*, A.K. Aziz ed., Academic Press, 1972.
- [11] P.-A. Raviart and J.-M. Thomas. Mixed Finite Element Methods for Second Order Elliptic Problems. *Lecture Notes in Mathematics*, Springer Verlag, 606: 292–315, 1977.
- [12] V. Ruas. Automatic generation of triangular finite element meshes. *Computer and Mathematics with Applications*, 5 (1979) 125–140.
- [13] V. Ruas and M.A. Silva Ramos. A Hermite finite element for Maxwell’s equations. *AIP Proceedings of ICNAAM*, T. Simos ed., Rhodes, Greece, 2016.
- [14] V. Ruas. A modified RT_0 method for flow in porous media confined in a curved region. To appear.
- [15] L. R. Scott. *Finite Element Techniques for Curved Boundaries*. PhD thesis, MIT, 1973.
- [16] G. Strang and G. Fix. *An Analysis of the Finite Element Method*. Prentice Hall, 1973.
- [17] O.C. Zienkiewicz. *The Finite Element Method in Engineering Science*. McGraw-Hill, 1971.